MOLECULAR & CELLULAR TOXICOLOGY, Vol. 2, No. 1, 7-10, March 2006

Linkage Disequilibrium (LD) Mapping and Tagging SNP Selection of C-Fos Induced Growth Factor (Figf) Gene in Korean Population

Sook Kim[1], Yeon-Kyung Yoo[1], Hye-Yoon Jang[1], Eun-Soon Shin[1], Eun-Young Cho[1], Eugene Kim[2], Junghyun Namkung[2], Jun-Mo Yang[2] & Jong-Eun Lee[1]

[1]DNA Link, Inc. Seodaemun-Gu Yonhee-Dong 15-1, Seoul, Korea. 120-110
[2]Department of Dermatology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
Correspondence and requests for materials should be addressed to J.-E. Lee (jonglee@dnalink.com)

Accepted 10 February 2006

Abstract

We performed comprehensive SNP validation and linkage disequilibrium (LD) analysis of the c-fos induced growth factor (Figf) gene in Korean population. Out of 32 SNPs, only 9 SNPs were polymorphic in Korean population. Validated SNPs formed a single extended haplotype block with strong LD through the entire length of the gene. Tagging SNP analysis picked only 2 SNPs to represent most of the genetic variation information of the Figf gene. Our results demonstrate the utility of LD block and tagging SNP analysis for an efficient way of performing a candidate gene based association study.

Keywords: SNP, FigF gene, Haplotype, Linkage disequilibrium

A SNP association study is one of the ways to evaluate genetic factors involved in pathophysiology of diseases. Such approaches have found many genetic polymorphisms involved in many diseases. Until recently, SNP-centric candidate gene approaches of SNP association studies usually included only a few SNPs as a target and were hoping to find association with them. However, many experiments based on this approach failed due to little genetic information that the target SNPs carry. Unless picking the causative SNPs for a specific phenotype of interest, it is very hard to expect that the selected SNPs show significant association with the diseases. Thus, haplotype approach which combines the information of multiple adjacent SNPs into relatively simpler haplotype information is gaining more popularity since it can provide more genetic information and provide solutions regarding sample sizes[1]. Haplotype blocks are generally associated with limited haplotype diversity, such that a few major haplotypes explain the majority of the diversity in a block[2,3]. Linkage disequilibrium (LD) mapping based haplotype approach has become a powerful and cost-effective method for performing genetic association studies, particularly in the search for genes causing complex diseases like allergies[4]. Since the SNPs forming a haplotype are often in high LD within each other, selecting tagging SNPs can improve genotyping efficiencies greatly[5-7]. LD based haplotype mapping and selection of tagging SNPs to identify genetic factors related with complex diseases are becoming a standard approach for SNP association studies.

c-fos induced growth factor (Figf), also known as vascular endothelial growth factor (VEGF-D) is thought to be involved in many biological pathways including lymph vessel formation. Recently, biological pathways involved in lymph vessel formation were suggested to play a role in allergic diseases like asthma and atopic dermatitis[8]. It is located on X Chromosome and spans about 39 kb. It is consisted with 7 exons and there are 96 SNPs reported onto dbSNP.

Atopic dermatitis is a typical allergic disease and the genetic components of the disease are supposed to be multigenic. To investigate the involvement of Figf in the pathophysiology of atopic dermatitis in Korean population, we performed linkage disequilibrium analysis and haplotype mapping of the gene in Koreans. To have comprehensive coverage of the gene, total of 32 SNPs from dbSNP were selected for validation in Korean population. SNPs from 10 kb upstream to 5 kb downstream of the coding sequences including the entire gene were searched from the dbSNP. This is to have at least 1 kb of target SNP throughout the genic region of the Figf and hoping to have minimum of 2 kb of marker interval after validation. The results of the validation of SNPs on Korean population are shown in Table 1. Only 9 SNPs out of 32 showed any frequency in Korean population and

Table 1. Summary of validation of SNPs in Koreans.

rs_number	SNP Name	Position	Obs_het	Exp_het	MAF	HWE p-value
6632528	-330G/A	15162055	0.341	0.479	0.398	0.092
6632521	IVS1+566G/A	15161070	0.341	0.479	0.398	0.092
12843793	IVS1+1721G/A	15159915	0	0	0	1
12011360	IVS1+2822T/G	15158814	0	0	0	1
12858444	IVS1+4074G/A	15157562	0.114	0.107	0.057	1
6629049	IVS1+5623C/A	15156013	0.341	0.456	0.352	0.153
16979874	IVS1+6742T/C	15154894	0	0	0	1
1894322	IVS1+7653T/C	15153983	0	0	0	1
6632499	IVS1+8848G/A	15152788	0	0	0	1
6632494	IVS1-9453G/A	15150551	0	0	0	1
13440564	IVS1-9223T/G	15150321	0	0	0	1
16979861	IVS1-7657G/A	15148755	0	0	0	1
6418686	IVS1-6565G/A	15147663	0.419	0.478	0.395	0.562
6632474	IVS1-5622G/A	15146720	0.429	0.472	0.381	0.726
6629038	IVS1-4781T/C	15145879	0	0	0	1
12850684	IVS1-3053T/A	15144151	0	0	0	1
12011065	L37L	15141078	0	0	0	1
6629030	IVS2+122T/C	15140766	0.409	0.474	0.386	0.499
16997023	IVS2-1894G/C	15137866	0	0	0	1
6653956	IVS2-1482T/C	15137454	0	0	0	1
12858267	IVS2-587G/A	15136559	0	0	0	1
17216420	IVS3+314G/A	15135468	0	0	0	1
17216413	IVS3-1334G/A	15134411	0	0	0	1
16979835	IVS4+707T/C	15132222	0	0	0	1
12397241	IVS4-674T/C	15131621	0	0	0	1
16979833	IVS5+275T/A	15130562	0	0	0	1
12855116	IVS5-2054T/G	15127192	0	0	0	1
6527518	IVS5-698C/A	15125836	0.432	0.5	0.486	0.565
5980150	IVS6-318G/A	15124356	0	0	0	1
7064042	2262T/C	15122708	0	0	0	1
16979830	3261G/A	15121709	0	0	0	1
5935957	4407G/A	15120563	0.432	0.494	0.443	0.546

their detailed information regarding minor allele frequency (MAF), heterozygosity, and so on are listed in the table.

SNPs with MAF of over 10% were selected for further analysis. Linkage disequilibrium (LD) analysis was performed with Haploview program and showed strong LD throughout the gene. Selected markers were evenly spaced and they showed strong LD pattern which is ideal for haplotype analysis. The results of LD mapping are shown on Figure 1. Genes with strong LD patterns like this are ideal for tagging SNP approach since the markers are under strong correlations to each other and can provide a chance to reduce the number of genotyping by selecting tagging SNPs. Tagging SNP selection were performed with Tagger program on the web site and total of 2 SNPs were selected as tagging SNPs with LD bin tagging approach. 2 selected tagging SNPs are shown with green boxes. Since the markers were under strong LD to each other, each tagging SNPs have alternative SNPs to replace them as shown on Figure 1. The markers with complete linkage with each other were marked with yellow lines. This feature is good for selecting best working genotyping assays since not every SNPs show same quality of genotyping results with any specific genotyping technology.

Discussion

As shown on the results, LD mapping and tagging SNP selection procedure can give researchers with comprehensive view of the genomic structure of the gene and show the most economic way of analyzing the gene without losing much of genetic information. We demonstrated that the association study investigation the involvement of Figf gene on atopic dermatitis can be investigated with only 2 SNPs instead of 32 SNPs that are reported on dbSNP. Also, any SNPs that are with complete linkage with the selected tagging SNPs can replace them providing better chance of getting the best genotyping assays for the study.

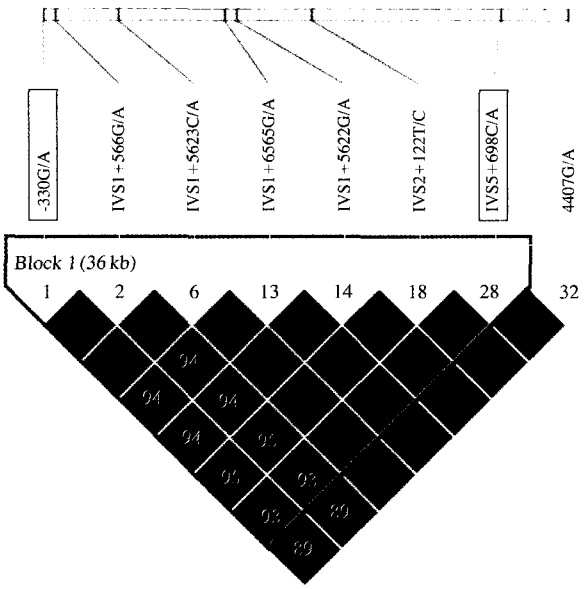

Fig. 1. LD block structure and tagging SNPs of Figf in Koreans. LD block pattern of the Figf gene in Koreans show strong LD pattern and formed one block. Relative location of the SNPs were shown on the top of the figure. 2 selected tagging SNPs are shown with green boxes. The markers with complete linkage with each other were marked with yellow lines.

Methods

DNA Samples and SNP Selection

The samples used for the study comprised of 90 Korean samples randomly selected from the cohort samples preserved in Korean National Institute of Health without family history of major diseases. SNP sites were selected from dbSNP database build 119 (http://www.ncbi.nlm.nih.gov/SNP) after applying repeat masking.

Genotyping

Multiplex SNP analyses were performed with the GenomeLab SNPstream genotyping platform (Beckman Coulter) and its accompanying SNPstream software suite as described by Demomme and Van Oene[9]. Briefly, the 12 pairs of PCR primers and single-base extension primers were optimally designed for each SNPs with Web-based software (http://www.autoprimer.com/ Beckman Coulter Inc. Fullerton, CA). PCR primers were designed to amplify a short stretch of genomic DNA (90-150 bp) that contains the target SNPs. The extension primers used to identify the SNPs contain unique tail sequences which is complementary to 1 of 12 unique tag oligonucleotides that are microarrayed at a specific location within each well of a 384-well microplate and short stretch of sequences that is complementary and precisely adjacent to the SNP. After a 12-plex PCR, the amplified fragments were treated with a mixture of exonuclease I and shrimp alkaline phosphatase (SNP-IT clean up kit, USB) to get rid of unincorporated PCR primers and dNTPs. The extension primers were added to the cleaned products and extended with fluorescent dye labeled nucleotide terminators and then spatially resolved by hybridization to the complementary tag oligonucleotides arrayed on the 384-well (SNPware Tag array, Beckman Coulter) microplates. The Tag array plates were analysed with a GenomeLab SNPstream array imager. The 12 individual SNPs were identified by their position and fluorescent color in each well according to the position of the tag oligonucleotides. Individual sample genotype data were generated on the basis of the relative fluorescent intensities for each SNP. All genotypes were reviewed and confirmed by an operator for final genotype calls. The average genotyping error rate was estimated at about 0.1% by routinely running duplicated sample wells in the plate.

LD Analysis

Only markers that were polymorphic in the Korean populations with MAF of over 10% were used for this and all the following analyses. Haplotype blocks were defined using HaploView[10]. Haplotypes and their frequencies in block regions were estimated using snphap (http://www-gene.cimr.cam.uk/clayton/software).

Tagging Analysis

Tagging SNPs were selected using Tagger (http://www.broad.mit.edu/mpg/tagger) with aggressive tagging and r^2 or haplotype $r^2 >= 0.80$, MAF cut-off = 5% and LOD threshold for multi-markers at 3.0. Tagging efficiency was defined as, n/n_h, where n_h is the number of tagging SNPs selected to cover the region and n is the total number of markers genotyped[11].

Acknowledgements

This work was funded by a grant (01-PJ3-PG6-01GN12-0001) of the 2001 Good Health R & D Project, Ministry of Health and Welfare, Republic of Korea. This study was also supported by a grant of the Korea Health 21 R & D Project, Ministry of Health & Welfare, Republic of Korea (A050558).

Reference

1. Lee, J.-E. *et al.* Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 195-204 (2005).
2. Daly, M.J. *et al.* High-resolution haplotype structure in the human genome. *Nat Genet.* **29**, 229-232 (2001).
3. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225-2229 (2002).
4. Morton, N.E. Linkage disequilibrium maps and association mapping. *J Clin Invest.* **115**, 1425-1430 (2005).
5. Johnson, G.C. *et al.* Haplotype tagging for the identification of common disease genes. *Nat Genet.* **29**, 233-237 (2001).
6. Stram, D.O. Tag SNP selection for association studies. *Genet Epidemiol.* **27**, 365-374 (2004).
7. Weale, M.E. *et al.* Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN-1A: implications for linkage-disequilibrium gene mapping. *Am. J. Hum. Genet.* **73**, 551-565 (2003).
8. Oliver, G. & Alitalo, K. The lymphatic vasculature: recent progress and paradigms. *Annu. Rev. Cell Dev. Biol.* **21**, 457-483 (2005).
9. Denomme, G.A. & Van Oene M. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* **45**, 660-666 (2005).
10. Barrett, J.C. *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
11. Ke, X. *et al.* The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet.* **13**, 577-588 (2004).